

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
29 April 2004 (29.04.2004)

PCT

(10) International Publication Number
WO 2004/036920 A1

(51) International Patent Classification: H04N 7/50, 7/36

(21) International Application Number:
PCT/IB2003/004522

(22) International Filing Date: 13 October 2003 (13.10.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
02292552.3 16 October 2002 (16.10.2002) EP

(71) Applicant (for all designated States except US): KONINKLIJKE PHILIPS ELECTRONICS N.V. [NL/NL];
Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(72) Inventors; and

(75) Inventors/Applicants (for US only): DRIEUX, Marc

[FR/FR]: 156 Boulevard Haussmann, F-75008 Paris (FR).
GROLIERE, Françoise [FR/FR]; 156 Boulevard Haussmann, F-75008 Paris (FR). SNOOK, Daniel [GB/FR];
156 Boulevard Haussmann, F-75008 Paris (FR).

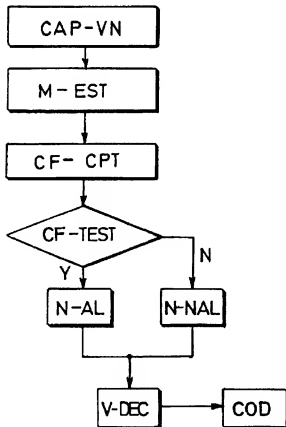
(74) Agent: LANDOUSY, Christian; Société Civile SPID,
156 Boulevard Haussmann, F-75008 Paris (FR).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,

[Continued on next page]

(54) Title: VIDEO ENCODING METHOD



(57) Abstract: The present invention relates to an encoding method processing video object planes (VOPs) of I, P or B type and including a dynamic allocation of said VOPs. According to said method, for each input frame a preliminary forward motion estimation is carried out between current and previous frames. Current and previous motion fields are then used to evaluate a coherence factor, expressed as connected to the sums of local differences within current and previous motion field. According to the value of the coherence factor with respect to an empirically determined threshold, a decision is finally taken on the type of the frame to be coded.

WO 2004/036920 A1



ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO

patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW). Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM). European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR). OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

Published:

- with international search report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

"VIDEO ENCODING METHOD"

FIELD OF THE INVENTION

5 The present invention generally relates to the field of object-based video coding, and more particularly to a method for encoding a sequence of video data according to a process in which several types of data are identified, said data consisting of so-called Video Object Planes (VOPs) that are either intra coded VOPs (I-VOPs), coded using information only from themselves, or predictive coded VOPs (P-VOPs), coded using a motion compensated
10 prediction from a past reference VOP, or bidirectionally predicted VOPs (B-VOPs), coded using a motion-compensated prediction from past and future reference VOPs.

BACKGROUND OF THE INVENTION

 The MPEG-4 visual standard provides technologies to view, access and manipulate objects (rather than pixels, in the case of the previous MPEG standards) in a large
15 range of bit rates, in a lot of application areas such as for instance : digital television, streaming video, mobile multimedia, games, etc. Said standard operates on video objects (VOs) defined by temporal and spatial information in the form of shape, motion and texture information, coded separately in the bitstream (these VOs are the entities that the user can access and manipulate).

 The MPEG-4 approach relies on a content-based visual data representation
20 of the successive scenes of a sequence, each scene being a composition of VOs with its intrinsic properties : shape, motion, texture. In addition to the concept of VO, the MPEG-4 standard introduces other ones like the Video Object Layer (each VO can be encoded either in a scalable or non-scalable form, depending on the application, represented by the video object layer, or VOL) and the Video Object Planes (VOPs) (= instances of VOs in time). It is assumed that each
25 frame of an input video sequence is segmented into a number of arbitrarily shaped image regions (the VOs), and that the shape, motion and texture information of the VOPs belonging to the same VO is encoded and transmitted into separate VOLs corresponding to specific temporal or spatial resolutions (which allows later to separately decode each VOP and leads to the required flexible manipulation of the video sequence).

30 The three types of frames processed by such a coding structure are the following : the I-VOPs, the P-VOPs and the B-VOPs. An I-VOP is an intra coded VOP : the coding operation uses information only from itself (it is the VOP that costs the greatest number of bits). A P-VOP is a predictive coded VOP, and the coding operation then uses a motion

- compensated prediction from a past reference VOP which can be either an I-VOP or another P-VOP (contrary to an I-VOP, only the difference between the current motion-compensated P-VOP and its reference is coded : thus, a P-VOP usually costs fewer bits than an I-VOP). A B-VOP is a VOP that is coded using a motion-compensated prediction from past and future
- 5 reference VOPs (I or P-VOPs), based on so-called forward and backward motion estimations respectively. A B-VOP cannot be a reference VOP and, like the P-VOP, only the difference between the current motion compensated B-VOP and its reference VOP is coded.

- Unfortunately, using said B-VOP prediction (also called interpolated or bi-directional mode) is not always a gain in term of compression. If the compression can sometimes
- 10 be improved by a factor of about 20 %, it can also in other cases be decreased by a drastic factor.

SUMMARY OF THE INVENTION

It is then an object of the invention to propose an encoding method using this B-VOP prediction only when it is efficient.

- To this end, the invention relates to an encoding method such as defined in
- 15 the introductory part of the description, said encoding method including a coding step of each VOP and, before said coding step, a motion estimation step performed between the current VOP and the previous one, said motion estimation step itself comprising a decision process concerning the type of VOP to be coded and based on the sub-steps of :
- carrying out a motion estimation between a VOP number N (VOP N) and
 - 20 the previous one (VOP N-1) ;
 - on the basis of said motion estimation, computing a so-called coherence factor, provided for quantifying the sequence motion ;
 - on the basis of a comparison of said coherence factor with a predetermined
 - threshold, taking a final decision on the type of the current VOP, said current VOP being a B-
 - 25 VOP or not according to the value of said coherence factor with respect to said threshold.

BRIEF DESCRIPTION OF DRAWINGS

The present invention will now be described, by way of example, with reference to the accompanying drawings in which Fig.1 illustrates the main steps of the encoding method according to the invention.

30 DETAILED DESCRIPTION OF THE INVENTION

An MPEG-4 encoder comprises several functional blocks, among which one or several memories, for outputting the VOPs in the transmitting order required by the standard.

For example, if the input sequence is I B B P B B P..., the output or transmission order will be I P B B P B B... (to enable backward prediction, the encoder must re-order the pictures from natural order to said transmission order so that the B-VOPs are transmitted after the past and future VOPs which are their references). Said encoder also comprises a motion estimator, for receiving the current VOP and the previous one (or reference VOP) and taking the decision of which kind of prediction will be implemented for the current VOP : no prediction for an I-VOP, forward prediction for a P-VOP, bi-directional prediction for a B-VOP.

As illustrated in Fig.1 showing the steps of the decision about the prediction within the motion estimator, the decision is, according to the invention, based on the implementation of the following steps. First, the current VOP (number : N) is captured (step CAP-VN). Then a motion estimation is carried out (step M-EST) between the VOP N and the previous VOP (number : N-1), and a factor named "coherence factor" is computed (step CF-CPT), in order to quantify the sequence motion, and compared (step CF-TEST) to a predetermined threshold. According to the result of the comparison (Y = lower than, or N = not lower than), the VOP N is allowed (N-AL), or not (N-NAL), to be a B-VOP. The final decision concerning the prediction mode is then taken (step V-DEC), and the coding step of the current VOP (= I-VOP, or P-VOP, or B-VOP) can take place (step COD).

Several expressions may be proposed for the coherence factor used in the comparison test, without limiting the scope of the invention. However, a preferred coherence factor may be for instance expressed as the ratio of the sum of absolute differences (SAD) between motion vectors of a macroblock (estimated in 16 x 16 pixels mode or 8 x 8 pixels mode) and its predecessor in the same VOP with the similar sum for the previous VOP (it is here recalled that for a macroblock of size k x k, the expression of the SAD is :

$$SAD = \sum_{i=0}^{bck} |A(i) - B(i)|$$

where B(i) and A(i) respectively designate the current macroblock considered and the macroblock in the reference VOP which matches the most in a search zone defined in said reference VOP).

CLAIMS :

1. A method for encoding a sequence of video data according to a process in which several types of data are identified, said data consisting of so-called Video Object Planes (VOPs) that are either intra coded VOPs (I-VOPs), coded using information only from themselves, or predictive coded VOPs (P-VOPs), coded using a motion compensated prediction from a past reference VOP, or bidirectionally predicted VOPs (B-VOPs), coded using a motion-compensated prediction from past and future reference VOPs, said encoding method including a coding step of each VOP and, before said coding step, a motion estimation step performed between the current VOP and the previous one, said motion estimation step itself comprising a decision process concerning the type of VOP to be coded and based on the sub-steps of :
 - carrying out a motion estimation between a VOP number N (VOP N) and the previous one (VOP N-1) ;
 - on the basis of said motion estimation, computing a so-called coherence factor, provided for quantifying the sequence motion ;
 - on the basis of a comparison of said coherence factor with a predetermined threshold, taking a final decision on the type of the current VOP, said current VOP being a B-VOP or not according to the value of said coherence factor with respect to said threshold.
2. An encoding method according to claim 1, in which said coherence factor is expressed as the ratio of the sum of absolute differences (SAD) between motion vectors resulting from said motion estimation, for a macroblock and its predecessor in the same VOP, with the similar sum for the previous VOP.

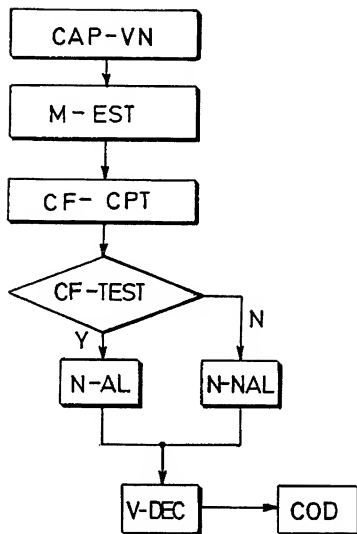


FIG.1

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 H04N/50 H04N7/36

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 6 307 886 B1 (WESTERMANN EDWARD F) 23 October 2001 (2001-10-23) abstract column 6, line 55 -column 7, line 10; figure 6	1,2
X	US 2002/037051 A1 (TAKENAKA YUJI) 28 March 2002 (2002-03-28) abstract paragraph '0065! - paragraph '0066!; claims 1-4; figures 1,5 --- -/-	1

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *Z* document member of the same patent family

Date of the actual completion of the international search

15 January 2004

Date of mailing of the international search report

22/01/2004

Name and mailing address of the ISA
European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax (+31-70) 340-3016

Authorized officer

Foglià, P

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>LEE L-W ET AL: "MOTION ORIENTED PICTURE INTERPOLATION WITH HUMAN PERCEPTUAL CONSIDERATION"</p> <p>JOURNAL OF VISUAL COMMUNICATION AND IMAGE REPRESENTATION, ACADEMIC PRESS, INC, US, vol. 4, no. 4, December 1993 (1993-12), pages 298-308, XP001086696</p> <p>ISSN: 1047-3203</p> <p>page 300, paragraph III -page 302; figures 2,3</p> <p>---</p>	2
A	<p>LEE J ET AL: "TEMPORALLY ADAPTIVE MOTION INTERPOLATION EXPLOITING TEMPORAL MASKING IN VISUAL PERCEPTION"</p> <p>IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE INC. NEW YORK, US, vol. 3, no. 5, 1 September 1994 (1994-09-01), pages 513-526, XP000476828</p> <p>ISSN: 1057-7149</p> <p>page 515, left-hand column, line 3</p> <p>-right-hand column, line 12</p> <p>-----</p>	1,2

Patent document cited in search report		Publication date		Patent family member(s)	Publication date
US 6307886	B1	23-10-2001	JP SG	11275585 A 74691 A1	08-10-1999 22-08-2000
US 2002037051	A1	28-03-2002	JP	2002101416 A	05-04-2002

BEST AVAILABLE COPY